Dossiers Personnels Ubiquitaires et Sécurisés

Nicolas Anciaux, Morgane Berthelot, Martine de La Blache, Luc Bouganim, Laurent Braconnier, Georges Gardarin, Philippe Kesmarszky, Sophie Lartigue, Jean-François Navarre, Philippe Pucheral, et al.

Dossiers Personnels Ubiquitaires et Sécurisés

Nicolas Anciaux[1], Morgane Berthelot[2], Martine De la Blache[3], Luc Bouganim[1], Laurent Braconnier[3], Georges Gardarin[5], Philippe Kesmarszky[6], Sophie Lartigue[7], Jean-François Navarre[3], Philippe Pucheral[1,5], Jean-Jacques Vandewalle[4]

(1) INRIA Rocquencourt
78153 Le Chesnay Cedex, France
<Fname.Lname>@inria.fr

(2) SANTEOS SA
Tour Manhattan - 5,6 Pl. de l'Iris
92926 Paris la Défense Cedex, France
<Fname.Lname>@santeos.com

(3) Yvelines district council
Hôtel du Département - 2 Pl. A. Mignot
78012 Versailles Cedex, France
<F.Lname>@cg78.fr

(4) Gemalto
6 Rue de la Verrerie
92190 Meudon,France
<Fname.Lname>@gemalto.com

(5) PRISM Laboratory
University of Versailles - 45 av. des Etats-Unis
78035 Versailles Cedex, France
<Fname.Lname>@prism.uvsq.fr

(6) ALDS (Association Locale de Développement Sanitaire)
25, av. des Aulnes - 78250 Meulan, France
kph@alds.org

(7) CoGITEY (Coordination Gérontologique Intercommunale du Territoire Est Yvelines)
6 av. Mal F. d'Esperey - 78004 Versailles, France
<F.Lname>@cogitey.com

1 Introduction

Les solutions existantes de partage de données (médicales, sociales, administratives, commerciales, professionnelles, etc.) sont classiquement basées sur une approche serveur. L'approche serveur apporte en effet des propriétés essentielles telles que : *complétude* (c.à.d, l'information est complète et à jour), *disponibilité* (l'information est accessible 7j/7 et 24h/24 via l'Internet), *usage* (l'information est organisée, facilement interrogeable et exploitable), *cohérence* (respect des contraintes d'intégrité, atomicité des mises à jour, isolation des traitements concurrents), *durabilité* (tolérance aux pannes) et *sécurité* (protection vis-à-vis des accès illégitimes).

Les solutions serveur souffrent cependant de deux carences. La première tient à l'impossibilité d'accéder aux données sans une connexion fiable, sécurisée, permanente et rapide au serveur, un ensemble de conditions difficile à réunir dans tous les environnements. La seconde réside dans la défiance des utilisateurs envers une gestion centralisée de leurs données personnelles. Cette défiance s'explique tout autant par l'absence de garantie de sécurité dès lors que les données quittent la zone sécurisée du serveur que par une perte de contrôle de l'utilisateur sur la façon dont ses données sont partagées et exploitées par différents acteurs.

Ce document décrit la conception d'une nouvelle forme de dossier personnel ubiquitaire et sécurisé et son expérimentation dans un contexte d'échange de données médico-sociales[1]. L'objectif est de palier les carences précitées, non pas en se substituant à l'approche centralisée classique, mais plutôt en la complétant de manière appropriée. L'approche proposée s'articule autour d'un nouveau composant matériel appelé ici SPT (Secure Portable Token), qui associe la sécurité intrinsèque d'une carte à puce à la capacité de stockage d'une clé USB (à terme plusieurs Giga-octets) et à l'universalité du protocole USB (lecture à partir de tout terminal équipé d'un port USB : station de travail, PC portable, assistant personnel, téléphone cellulaire, etc.). Notamment, un SPT peut héberger un réplica de tout ou partie d'un dossier de données personnelles géré par le serveur pour permettre des traitements déconnectés. Par ailleurs, les capacités sécuritaires du SPT peuvent être exploitées pour mettre en œuvre de nouveaux schémas de partage des données, fortement sécurisés et directement contrôlables par l'utilisateur.

[1] Cette initiative est financée pour partie par l'Agence Nationale de la Recherche dans le cadre du projet RNTL-PlugDB et pour partie par le Conseil Général des Yvelines dans le cadre du projet DMSP.

Plus precisément, le problème attaqué revêt quatre dimensions. Les trois premières sont relatives aux appréhensions introduites par la centralisation des informations personnelles, la dernière découle directement du problème de l'accès déconnecté.

1. *Expression du consentement* : rendre à l'individu la possibilité de contrôler directement la divulgation d'informations particulièrement sensibles le concernant, comme s'il détenait ces données à domicile, dans un dossier papier.
2. *Conservation des données* : rendre à l'individu la maîtrise de la durée de conservation de ses données, et ainsi lui permettre de détruire définitivement certaines données (droit à l'oubli).
3. *Continuité de la sécurité :* garantir le même niveau de sécurité quel que soit l'endroit où les données sont hébergées (serveur ou terminal) et quelle que soit la façon dont elles sont manipulées (accès connecté ou non).
4. *Accès déconnecté :* permettre des accès déconnectés aux données tout en garantissant une cohérence à terme de ces données.

Au-delà de la résolution de ces quatre dimensions, ce projet s'attache à l'intégration de l'équipement nomade dans une infrastructure globale. Le principe de dossier portable sécurisé ne peut en effet se concevoir en isolation. Le contenu d'un dossier portable a vocation à être intégré dans un système d'information global, qu'il l'alimente ou qu'il soit alimenté par ce dernier. De plus, tout équipement portable présente une vulnérabilité intrinsèque face aux risques de perte, de vol ou de destruction, imposant une réplication des données sur un serveur pour en assurer la durabilité et la disponibilité.

Le document est organisé comme suit. La section 2 détaille l'analyse des besoins qui a permis de fixer avec précision les dimensions du problème, préalable indispensable à la spécification de l'architecture fonctionnelle proposée. La section 3 introduit l'ensemble des composants de l'architecture, leur rôle respectif et leurs interactions. La section 4 conclut cet article.

2 Analyse des besoins et dimensions du problème

2.1 Perception des utilisateurs et législation

Nombre d'analyses ont été publiées sur les risques associés à l'accumulation de données à caractère personnel dans des bases de données et sur la perception des individus vis-à-vis de la préservation de leur vie privée. Il ressort de ces analyses qu'une forte proportion d'individus (de 70% à 80% selon les pays) a la sensation d'avoir perdu tout contrôle sur l'usage qui est fait de leurs données personnelles par des sociétés commerciales [IBM]. De même, aux USA, 70% des patients pensent que leurs données de santé sont mal protégées face aux attaques et qu'elles sont accédées sans qu'ils en aient connaissance, arrivant même à la conclusion (pour 48% des patients) que le risque de violation de la vie privée l'emporte sur le bénéfice escompté en terme de suivi médical, tout en reconnaissant ce bénéfice [West05].

La législation relative à la protection des données à caractère personnel [EuD95, HIPAA] introduit des garde-fous importants. Par exemple, la directive européenne 95/46/EC [EuD95] édicte les principes de *Licéité, Finalité, Exactitude, Droit d'accès, Information et consentement, Non discrimination, Sécurité,* et *Dérogation*. Malheureusement, ces principes généraux sont difficiles à traduire en termes informatiques (par exemple, la notion de consentement est sujette à de multiples interprétations). Ces principes sont également difficiles à garantir dans un contexte de forte criminalité informatique

[CSI06]. Les travaux sur les SGBD Hippocratiques [AKSX02] vont dans le sens d'une meilleure prise en compte des principes législatifs dès la conception des noyaux de SGBD. Cependant, un long chemin reste à parcourir. Par exemple, il a été montré qu'une opération aussi simple que la destruction d'une donnée dans une base, sans laquelle le principe de conservation limitée perd son sens, n'est supportée de façon effective par aucun SGBD commercial [MLS07]. En effet, des réminiscences des données détruites apparaissent à de multiples endroits (index, journaux de reprise après panne, journaux d'audit, etc).

Le projet n'a pas l'ambition d'apporter une réponse globale à ces problèmes dont chacun peut ouvrir une problématique de recherche à long terme. Par contre, des solutions convaincantes peuvent être apportées à des cas particuliers d'usage dont l'importance pratique est très significative et qui peuvent ouvrir la voie à des généralisations importantes. C'est dans cet esprit que la phase de spécification du projet a été conduite.

L'étude a démarrée par l'analyse des besoins concrets rencontrés sur le terrain par les professionnels de santé et les professionnels de l'action sociale au sein des deux coordinations gérontologiques (l'ALDS et COGITEY) participant au projet. Cette analyse a ensuite été généralisée à l'accès aux données médicales dans un contexte élargi et une réflexion a également été menée concernant l'accès à des données personnelles sensibles autre que médicales et sociales. Cette analyse a permis de préciser les dimensions du problème que se propose d'attaquer le projet. Sur cette base, de nouveaux scénarios d'échange de données sensibles rendus possibles par la technologie introduite ont été établis. Les sous-sections suivantes relatent les étapes de cette analyse.

2.2 Partage de données au sein d'une coordination gérontologique

Le vieillissement de la population rend crucial le suivi sanitaire des personnes âgées à domicile. Dans ce contexte, des informations personnelles d'ordre médicales et sociales doivent être échangées entre acteurs (médecins, infirmières, assistances sociales, aides ménagères, entourage) ayant des droits différents sur les données. Il est indispensable que ces données soient accessibles au chevet du patient, sans pouvoir tabler sur une connexion permanente et rapide à un serveur externe, rare chez les personnes âgées.

Dans cet objectif, L'ALDS a déjà mis au point un « Dossier Médical Commun » au format papier, afin de permettre aux professionnels et intervenants du secteur médico-social de consigner les faits importants du suivi des personnes âgées. Ce dossier papier est conservé par le patient à domicile et le suit dans ses éventuels déplacements auprès des différents intervenants de son circuit médico-social. Chaque intervenant dispose, dans ce dossier, d'un espace privilégié dans lequel consigner les faits marquants survenus et devant être partagés avec les autres intervenants. Ainsi, les aide-ménagères, Aides-soignants, Assistantes sociales, Auxiliaires de vie, IDE, Kinésithérapeutes et médecins, d'une part, peuvent consulter le dossier et utiliser une feuille « tableau de bord » pour porter toute indication significative. Les médecins – hospitaliers ou spécialistes, peuvent quant à eux utiliser la feuille qui leur est réservée « séjours hospitaliers et consultations spécialisées » pour consigner les informations importantes.

Les différents intervenants du circuit médico-social du patient disposent également, individuellement, en règle générale, de leur propre outil de production : logiciel de cabinet médical, logiciel de service hospitalier, logiciel infirmier, logiciel de coordinations gérontologiques, logiciel du réseau Emile, etc.

Si elle a prouvé son efficacité au quotidien, l'utilisation de ce dossier papier se heurte à des difficultés majeures :

- Absence totale de confidentialité puisque tous les intervenants (médecins, infirmières, assistantes sociales, aides à domicile, entourage) ont accès à l'intégralité du dossier. L'ALDS doit pourtant faire face quotidiennement à des situations humainement complexes, comme par exemple : un patient se sachant en fin de vie mais ne voulant pas le dévoiler pour des raisons humaines et/ou financières (risque de détournement d'héritage) ; un patient désirant cacher une addiction (e.g, alcoolisme) ; un patient désirant cacher une pathologie particulière à ses proches, du fait de sa gravité (e.g., cancer) ou de son caractère dégradant (e.g., incontinence).
- Impossibilité d'accéder au dossier à distance. Par voie de conséquence, le dossier n'est pas systématiquement à jour, générant un suivi moins précis et une redondance de soins. La prise de décision distante (e.g, praticien interrogé au téléphone) ne peut pas non plus s'appuyer efficacement sur ce dossier.
- Ressaisies ultérieures des données intégrées à ce dossier dans de multiples applications informatiques avec le risque d'erreur et la perte de temps qui s'ensuit.

La solution envisagée passe naturellement par une dématérialisation du « Dossier Médical Commun » papier. Pour autant, la dématérialisation du dossier à elle seule ne répond pas à l'ensemble des manques identifiés ci-dessus. Ainsi, centraliser une copie électronique du dossier sur un serveur permet des accès à distance, la mise en place d'une politique de contrôle d'accès, la synchronisation avec d'autres outils de production. Mais la question des accès au dossier en mode déconnecté reste posée. Par ailleurs, la centralisation du dossier introduit de nouvelles craintes au regard de la confidentialité. Comment expliquer une politique de contrôle d'accès à une personne âgée ? Comment la persuader que tous les accès à son dossier seront parfaitement légitimes alors que ces accès sont eux-mêmes « dématérialisés » et difficilement contrôlables par le patient ? Les situations humaines complexes rencontrées par l'ALDS rendent la perception de la perte de contrôle des données médico-sociales particulièrement aiguë.

2.3 Partage général de données de santé

La situation rencontrée dans les coordinations gérontologiques gérées par l'ALDS et COGITEY est représentative des problèmes liés à l'échange de données médicales dans un contexte de réseau de soins spécialisé. Dans ce contexte particulier, le contenu des dossiers et la liste des intervenants habilités à interagir avec ce dernier sont connus. Le problème est encore plus complexe lorsqu'il s'agit d'organiser un dossier médical général amené à suivre le patient au cours de toute son histoire, et ce quelles que soient les facettes de cette histoire.

Au cours des dix dernières années, de nombreux pays ont lancé des programmes ambitieux de dossiers médicaux informatisés, dénommés usuellement systèmes EHR (Electronic Healthcare Record) dans l'objectif d'accroître la qualité des soins tout en diminuant leur coût. L'intérêt de centraliser l'information dans un système EHR est multiple : *complétude* (information complète et à jour), *disponibilité* (information accessible 24h sur 24h via un réseau), *usage* (information organisée, décrite, facilement interrogeable), *cohérence* (respect des contraintes d'intégrité, atomicité et isolation des transactions), *durabilité* (tolérance aux pannes) et *sécurité* (protection contre les accès illégaux).

Cependant, les craintes liées à la protection des données, au droit à l'oubli mais aussi aux modifications des relations patients-médecins sont à l'origine de réactions, parfois virulentes, des

professionnels de santé et des associations d'usagers et de malades [IC03]. Nous identifions quatre raisons principales à cette défiance :

- *Difficulté d'expression du consentement :* Le patient est censé donner son consentement sur une politique de contrôle d'accès régulant « qui » (individu ou rôle) peut accéder à quelle partie de son dossier. Même avec l'aide d'un praticien, il parait difficile d'obtenir un consentement éclairé. Ceci est dû à un nombre important d'acteurs pouvant potentiellement accéder légitimement à un dossier (plus de 150 en cas d'hospitalisation, selon le Los Angeles Times [LAT06]), à la complexité de l'information médicale et à la difficulté intrinsèque de préciser a priori quelle donnée élémentaire peut révéler quelle pathologie. En conséquence, le patient est souvent amené à consentir à la mise en place d'une politique de contrôle d'accès qu'il ne maîtrise pas. Des outils d'audit sont prévus pour permettre à un patient un contrôle a posteriori mais la complexité d'interprétation du journal d'audit augmente avec sa taille [ABF+04], rendant illusoire un contrôle efficace par le patient lui-même.
- *Conservation mal bornée des données :* la conservation des données bornée dans le temps est un principe fondateur de la législation protégeant les données personnelles [EuD95]. Toute donnée doit ainsi être détruite après que la durée de rétention légale ait expirée. Cependant, cette durée doit permettre l'accomplissement de l'objectif pour lequel la donnée est collectée. Elle peut de ce fait être particulièrement longue, par exemple de 40 à 50 ans [ACA05], et n'est pas toujours justifiée du fait que la satisfaction de l'objectif lui-même peut être difficile à apprécier. Par ailleurs, nous avons déjà évoquée la difficulté technique pour un SGBD de détruire physiquement une donnée [MLS07]. Ceci renforce le sentiment du patient que tous les événements de son histoire sont enregistrés à jamais. L'effet de bord est que le patient peut avoir tendance à écarter de son dossier une donnée (synonyme de dossier incomplet et de suivi des soins imparfait) s'il a un doute sur la sensibilité de cette donnée.
- *Absence de sécurité en dehors de la sphère du serveur :* les données médicales sont sujettes à extraction et peuvent être hébergées sur un terminal vulnérable (par exemple, celui d'un médecin ou du patient). Il s'agit même de la situation par défaut lorsque les données doivent être utilisées hors connexion (par exemple au chevet du patient). Cette situation risque de perdurer longtemps, du moins jusqu'à ce que l'intégralité du territoire soit couverte par un réseau sécurisé, rapide, fiable et gratuit. Malheureusement, un terminal est beaucoup plus vulnérable que le serveur aux virus, chevaux de Troie et autres espions, introduisant une importante brèche de sécurité dans l'architecture.
- *Pas d'accès déconnecté au dossier :* les systèmes EHR ont été conçus pour un usage en ligne. Ceci constitue une barrière importante pour une large catégorie de patients (ex : personnes âgées, nécessiteux), le pré-requis pour consulter son dossier étant d'utiliser un terminal mis à disposition dans un lieu public ou de posséder un PC, de maîtriser son usage (tâches d'administration comprises) et de payer pour une connexion internet. Dans le cas contraire, un médecin prodiguant des soins à domicile devra préalablement à sa visite télécharger du serveur toute donnée susceptible d'être utile à son diagnostic (une tâche fastidieuse et conduisant au trou de sécurité mentionné ci-dessus.

2.4 Partage d'autres types de données personnelles

La liste des données à caractère personnel susceptibles de subir des traitements informatisés est longue et la variété de ces mêmes traitements importante. Notre projet s'intéresse aux situations dans

lesquelles des informations personnelles sont centralisées et l'utilisateur concerné est directement partie prenante dans la façon dont l'échange d'information s'organise.

Nous pensons qu'à chaque fois que des informations personnelles sont centralisées et que la personne concernée doit exprimer son consentement, les deux premières difficultés mentionnées dans la section 2.3 apparaissent (*Difficulté d'expression du consentement* et *Conservation mal bornée des données)*. Les deux autres difficultés mentionnées (*Absence de sécurité en dehors de la sphère du serveur* et *Pas d'accès déconnecté au dossier*) sont présentes ou non selon le contexte applicatif.

A titre d'exemple, considérons un environnement virtuel, usuellement dénommé VHE (Virtual Home Environnement), permettant à un utilisateur de retrouver son environnement (mail, bookmarks, cookies, agenda, carnet d'adresses, etc) quel que soit le terminal auquel il se connecte. Le contenu du VHE peut être géré sur un serveur, typiquement chez un fournisseur de services. Le VHE d'un utilisateur peut être partagé entre plusieurs personnes (ex : agenda partagé) et surtout entre plusieurs applications (éditeurs, navigateurs, logiciels de courrier, etc). L'expression du consentement se pose vis-à-vis des applications, d'où l'apparition de normes telles que P3P du W3C [CLM+02] dont la mise en œuvre s'avère moins simple qu'espéré initialement. La question de la durée de conservation des données se pose également envers le fournisseur de services hébergeant les données. L'accès au contenu du VHE à partir d'un terminal non sécurisé est bien entendu un problème important, de même que l'accès à un VHE en mode déconnecté (si l'on veut une continuité de service hors connexion).

Le contexte d'intelligence ambiante fournit un autre exemple. Considérons un environnement de télésurveillance destiné à prévenir des situations à risque, par exemple chez des personnes dépendantes (ce contexte n'est pas sans rapport avec celui de la coordination gérontologique même si le problème à traiter est différent). Des capteurs accumulent des données sur l'environnement de la personne et l'activité de cette dernière. Ces données sont ensuite accédées soit par des applications dont l'objectif est de déclancher une alerte, soit par des personnes de l'entourage ou des référents assurant un suivi à distance. On comprend aisément que les difficultés mentionnées préalablement apparaissent dans ce scénario, à l'exception du traitement en mode déconnecté qui n'a pas de sens dans ce contexte.

2.5 Dimensions du problème

En conclusion, le projet considère un contexte dans lequel :

- des données personnelles sont accumulées sur un serveur,
- l'individu concerné est partie prenante dans la façon dont ses données sont partagées.

Les dimensions du problème adressées par le projet sont les suivantes.

1. *Expression du consentement* : rendre à l'individu la possibilité d'établir un contrôle strict, compréhensible et observable sur la façon dont ses données sensibles sont échangées.
2. *Conservation des données* : rendre à l'individu la maîtrise de la durée de conservation de ses données, en distinguant les notions de conservation et de durabilité.
3. *Continuité de la sécurité :* garantir le même niveau de sécurité quel que soit l'endroit où les données sont hébergées (serveur ou terminal) et quelle que soit la façon dont elles sont manipulées (accès connecté ou non).
4. *Accès déconnecté :* permettre des accès déconnectés aux données tout en garantissant une cohérence à terme de ces données.

2.6 Principes de base

Sans rentrer dans la description technique de l'architecture et des challenges associés, nous illustrons ci-dessous comment la technologie proposée peut répondre au problème posé. Le principe est le suivant. On suppose que chaque utilisateur U possède (1) un dossier de données personnelles géré par un serveur central et (2) un SPT personnalisé. Ce SPT contient le certificat de U lui permettant une authentification forte auprès du serveur lorsqu'il accède à son dossier en mode connecté. Ce SPT contient également un replica du dossier personnel de U ainsi que des composants logiciels embarqués (notamment serveur Web et SGBD) donnant accès au dossier, en mode déconnecté, à partir de tout terminal équipé d'un port USB et d'un navigateur Web.

Pour que U puisse échanger des données avec des partenaires P, ceux-ci doivent également être équipés d'un SPT personnalisé. Le SPT d'un partenaire P est similaire à celui de U du point de vue logiciel et matériel, mais son rôle dans les interactions avec le dossier de U est particulier. Le SPT de P contient le certificat de P lui permettant une authentification forte auprès du serveur, qu'il soit central ou embarqué, lorsqu'il accède au dossier de U. Dans les deux cas, P subit la politique de contrôle d'accès fixée par U, identique sur le serveur central et le serveur embarqué.

Si P accède à des données de U et décide de les répliquer sur son terminal, par exemple pour y accéder en mode déconnecté sans la présence de U, ces données seront chiffrées avec une clé connue du SPT de P uniquement. Quand P interroge ces données, le SGBD embarqué dans son SPT y accède et les déchiffre pour répondre à la requête. Un principe similaire permet à U de déposer des données chiffrées sur le serveur central à destination de P qui pourra alors y accéder grâce au SGBD embarqué dans son SPT. La différence pour U entre déposer des données en clair ou en format chiffré sur le serveur central est la suivante. Dans le premier cas (données en clair), le partage des données est contrôlé par la politique de contrôle d'accès à laquelle U a consenti (avec les réserves mentionnées préalablement). Dans le second cas (données chiffrées), la politique de contrôle d'accès est doublée d'une obligation de partage physique de clés de chiffrement, partage qui ne peut s'organiser que nominativement et sous le contrôle total de U.

Pour organiser le partage, l'utilisateur U peut ainsi choisir différents statuts pour ses données :

- *Données régulières :* Les données dites régulières sont répliquées sur le serveur central et le serveur embarqué, protégées par la même politique de contrôle d'accès. La motivation de répliquer des données régulières dans le dossier embarqué est d'en assurer la disponibilité en mode déconnecté.
- *Données secrètes :* les données dites secrètes sont exclusivement stockées sur le SPT de U. Au même titre que pour un dossier papier, U garde la liberté de donner accès à son SPT, et donc aux données secrètes, au partenaire P qu'il a physiquement en face de lui. Il a la garantie que personne ne peut accéder à ces données sans sa connaissance. La durabilité des données secrètes reste par contre à la charge de U.
- *Données secrètes durables :* il s'agit de données secrètes répliquées sur le serveur central dans un format chiffré par des clés de chiffrement connues exclusivement du SPT de U. Le serveur en assure la durabilité au même titre que pour des données régulières mais U garde la garantie que personne ne peut accéder à ces données sans détenir le SPT de U. Seule la durabilité des clés de chiffrement reste à la charge de U (plusieurs solutions simples existent à ce problème).
- *Données restreintes :* il s'agit de données que U souhaite partager de façon exclusive avec un petit cercle de partenaires P de confiance, avec la garantie que personne d'autre ne peut accéder

à ces données. Pour ce faire, U dépose sur le serveur central, via son SPT, des données chiffrées dont les clés sont connues exclusivement par les SPT du cercle de confiance.

A tout moment, U garde la possibilité de changer le statut de ses données. S'il est possible d'abaisser le niveau de sécurité d'une donnée pour faciliter son partage, le processus inverse est plus incertain. Par exemple, une donnée régulière que l'on rendrait secrète aura déjà traversé un état partagé, et donc aura pu être consultée, copiée, etc. Au-delà de la mise en œuvre d'un consentement mieux contrôlé par l'usager, ces différents statuts permettent également de contrôler la durée de conservation des données, du moins tant qu'elles n'ont pas été placées dans l'état de données régulières.

Enfin, la réplication de données entre serveur central et serveur embarqué pose le délicat problème de la synchronisation. Lorsque les deux serveurs se trouvent connectés ensemble, la synchronisation se déroule de façon traditionnelle. Cependant, cette situation peut ne jamais se produire (ex : un patient U ne sortant jamais de son domicile). Dans ce cas particulier, un protocole de synchronisation par procuration est mis en place, dans lequel les SPT de participants P rentrés en contact avec U vont servir de réseau en embarquant des messages chiffrés déchiffrables uniquement par U et le serveur central.

2.7 Scénarios

A titre illustratif, nous détaillons certains des principes ci-dessus à travers des scénarios concrets dans le contexte médico-social, schématisés sur la figure 1. Nous illustrons notamment la synchronisation entre le serveur central et le SPT, l'approvisionnement du dossier en données (mêmes non régulières), le partage avec le cercle de confiance, et la sécurité offerte aux composants non surs (comme un terminal).

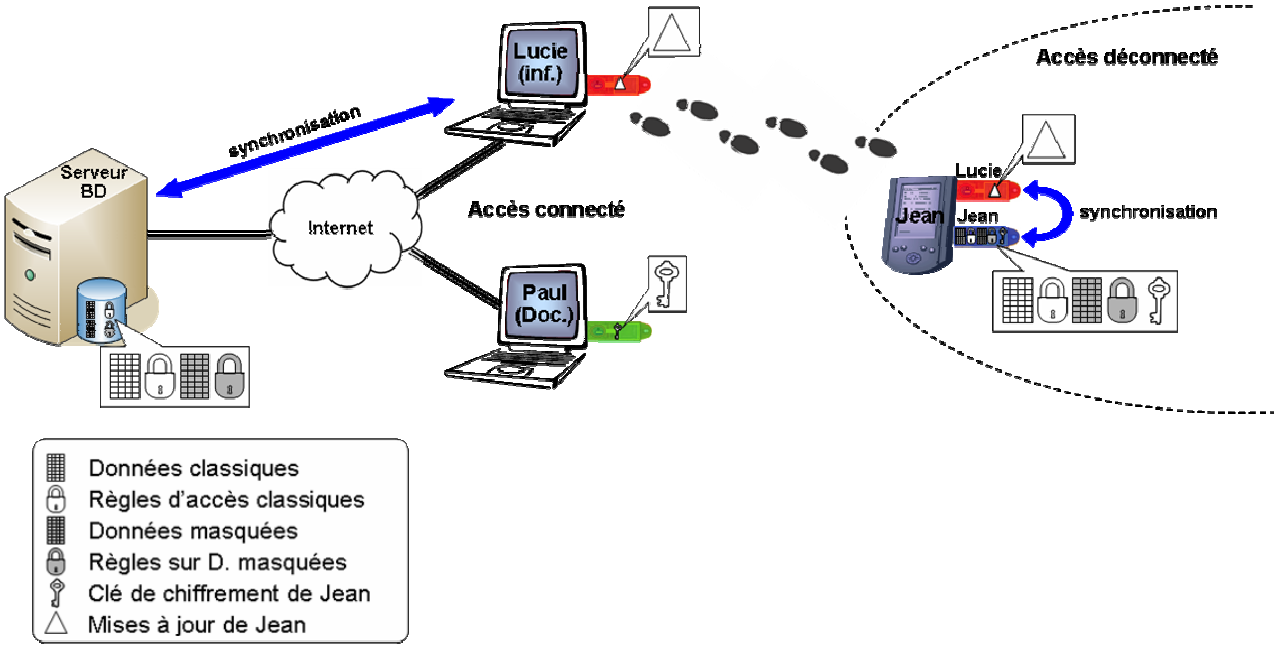

Figure 1 : échange de données médico-sociales utilisant les techniques proposées.

2.7.1 Intégration d'un nouvel événement dans le dossier au chevet du patient

Avant de visiter ses patients, le professionnel médico-social P se connecte au serveur central (authentification forte grâce au certificat contenu dans son SPT) et récupère les nouveaux éléments intégrés dans les dossiers de ses patients. Ces données sont chargées automatiquement dans le SPT de P sans que ce dernier n'y ait accès. Le SPT de P est utilisé comme canal de communication avec les SPT des patients, canal par lequel peuvent transiter des données auxquelles P n'a pas droit.

Arrivé au domicile du patient U, le professionnel P connecte son SPT et celui de U dans un terminal mobile. Il s'authentifie auprès de son SPT par un PIN code, puis son SPT s'authentifie auprès du SPT du patient grâce au certificat de P (il y a donc authentification forte de P auprès du serveur embarqué). Le SPT de U intègre les mises à jour le concernant en provenance du serveur central. P interagit ensuite avec le dossier de U en local via un navigateur Web. Il peut faire toutes consultations et mises à jour autorisées par la politique de contrôle d'accès. Les mises à jour sont signées grâce au certificat de P. En fin de visite, les mises à jour présentes dans le SPT de U et non présentes sur le serveur central sont chargées dans le SPT de P afin d'être communiquées au serveur central (le SPT de P sert donc de canal de communication bidirectionnel). Plus précisément, la synchronisation entre le serveur embarqué et le serveur central ne concerne que les données régulières. Les données secrètes ne sortent jamais du SPT de U. Les données secrètes durables et les données restreintes sont quant à elles chargées dans le SPT de P mais chiffrées avec des clés connues uniquement du SPT de U.

2.7.2 Intégration d'un nouvel événement dans le dossier hors de la présence du patient

Il peut s'agir par exemple de l'intégration d'un résultat d'analyse biologique. Ce résultat se réfère à une donnée déjà existante (une prescription) et hérite donc du statut de cette donnée. S'il s'agit d'un statut de donnée régulière, le résultat est classiquement intégré au dossier sur le serveur central et sera reporté sur le dossier embarqué à la prochaine synchronisation. Dans les autres cas, le résultat doit être chiffré avec une clé secrète k (choisie par le cabinet d'analyse), elle-même chiffrée avec la clé publique du patient concerné avant d'être intégré au serveur[2]. Lors d'une prochaine synchronisation, le SPT du patient accédera à ce résultat et le chiffrera avec les bonnes clés, en fonction de son statut (i.e., secret durable ou restreint) ou le supprimera du serveur (si le statut est secret). Rappelons que seul le patient a la possibilité de chiffrer ses données avec une clé secrète connue de son seul SPT.

2.7.3 Echange de données restreintes dans un cercle de confiance

Le patient U a décidé de partager une donnée D exclusivement avec son médecin référent P. Pour ce faire, le SPT de U chiffre D avec une clé secrète k et stocke sur le serveur central la version chiffrée de D, ainsi que la clé k, elle-même chiffrée avec la clé publique de P. Le médecin P peut être amené à consulter D sans la présence du patient. Le SPT de P récupère et déchiffre alors la clé k (grâce à la clé privée de P), ce qui lui permet ensuite de déchiffrer D.

2.7.4 Interrogation du dossier centralisé

Le professionnel P se connecte au serveur central (authentification forte grâce au certificat contenu dans son SPT) et récupère le résultat de sa requête. S'il décide de stocker ce résultat sur son terminal, ce dernier sera stocké chiffré avec une clé connue du SPT de P seul afin de se prémunir contre une attaque du terminal de P.

---

[2] Ce même mécanisme peut être utilisé pour donner l'accès à ces résultats au prescripteur.

3 Architecture fonctionnelle

La figure 2 présente l'architecture fonctionnelle du projet en distinguant les aspects infrastructure, données et logiciel. Au niveau logiciel, les composants grisés correspondent à des éléments logiciels existant dans le commerce (serveur Web, navigateur, SGBD relationnel, système d'exploitation) alors que les autres éléments sont spécifiquement développés dans le cadre du projet.

3.1 Infrastructure

La partie Infrastructure de la figure (de gauche à droite) représente le **serveur central** hébergeant les dossiers personnels, un utilisateur (P) connecté via un **terminal** ayant potentiellement rapatrié localement une partie de la base de données (pour utilisation en mode déconnecté), ainsi qu'un autre utilisateur (U) synchronisant son **SPT** sur un **PDA** grâce aux mises à jour transmises par le SPT de P.

L'accès de l'utilisateur P au serveur se fera via un **canal sécurisé** de type SSL / TLS qui garantira la confidentialité des échanges. L'utilisateur P et le serveur s'authentifieront de manière mutuelle, le serveur en présentant au terminal de P son certificat et P en présentant au serveur son propre certificat (ex : certificats de type GIP-CPS dans un contexte médical). Lors de cette présentation, les deux parties s'assureront que les deux certificats sont valides et non révoqués. L'ensemble de ces opérations se fera de manière transparente pour l'utilisateur. La seule opération active de l'utilisateur sera la saisie du code PIN sur son SPT.

Le serveur exposera une interface de type **serveur Web** donnant la possibilité de consulter, créer, ou synchroniser des données dans un dossier personnel via un navigateur Web. Toutes ces opérations seront effectuées via le canal SSL / TLS établi au préalable, pour garantir leur confidentialité.

Cette infrastructure repose sur une architecture matérielle traditionnelle à l'exception du **SPT**. Le SPT est une plate-forme matérielle dotée d'un microcontrôleur sécurisé de type carte à puce et d'une mémoire de stockage de masse de type Flash NAND pouvant atteindre plusieurs centaines de mega-octets, voire à terme plusieurs giga-octets. Bien qu'extérieure à la puce sécurisée et ne jouissant pas de la protection matérielle de celle-ci, la mémoire de masse de type FLASH est sous le contrôle de la puce : tous les accès en écriture et lecture à la mémoire de masse sont effectués par la puce sécurisée. Le SPT communique avec le monde extérieur via une interface USB.

3.2 Données

La partie Données de la figure montre le type et le format de chiffrement des données stockées dans chaque environnement, en fonction du statut de ces données. Le serveur central est amené à stocker des données d'administration telles que les **certificats** des utilisateurs et leur **clé publique**. Le serveur stocke également des **clés secrètes de chiffrement k**, elles-mêmes chiffrées avec les clés publiques des utilisateurs participant à un même cercle de confiance. Enfin, le serveur central stocke, pour chaque dossier personnel d'un utilisateur U, les **données régulières** (protégées par les moyens propres à l'hébergeur de données), les **données secrètes durables** chiffrées avec une (ou des) clé secrète k_U connue du SPT de U uniquement, les **données restreintes** chiffrées avec une (ou des) clé secrète k. Le terminal d'un utilisateur P peut **héberger des données** pour les rendre accessible en mode déconnecté. Elles sont alors chiffrées avec une (ou des) clé secrète k_P connue uniquement du SPT de P. Enfin, un SPT contient différent types de données suivant qu'il est utilisé comme SPT de l'utilisateur U pour gérer le dossier de U ou bien comme SPT de l'utilisateur P pour accéder à des données du dossier de U (les deux rôles pouvant être concomitants). Dans le premier cas (gestion par l'utilisateur de son propre dossier), le SPT contient des données administratives telles que le **certificat**

de l'utilisateur, sa **clé privée** (correspondant à sa clé publique présente sur le serveur central), un ensemble de **clés secrètes k_U** destinées au chiffrement des données secrètes durables sur le serveur central, un ensemble de **clés secrètes k** destinées au chiffrement des données restreintes. Ces données administratives sont stockées dans la Flash NOR interne au microcontrôleur sécurisé et donc protégée des attaques physiques. Les données du dossier, quel que soit leur statut, sont stockées dans la Flash NAND externe (non sécurisée matériellement) et chiffrées avec des clés secrètes de type k_U. Pour gérer le second cas d'usage (interaction de l'utilisateur avec un dossier autre que le sien), le SPT doit pouvoir servir de canal de communication entre le serveur central et le dossier avec lequel il interagit. Pour cela, le SPT est amené à transporter dans la Flash NAND des **données en transit** (appelées Deltas sur la figure) chiffrées avec une (ou des) clé secrète k_D, elle-même intégrée aux données administratives du SPT.

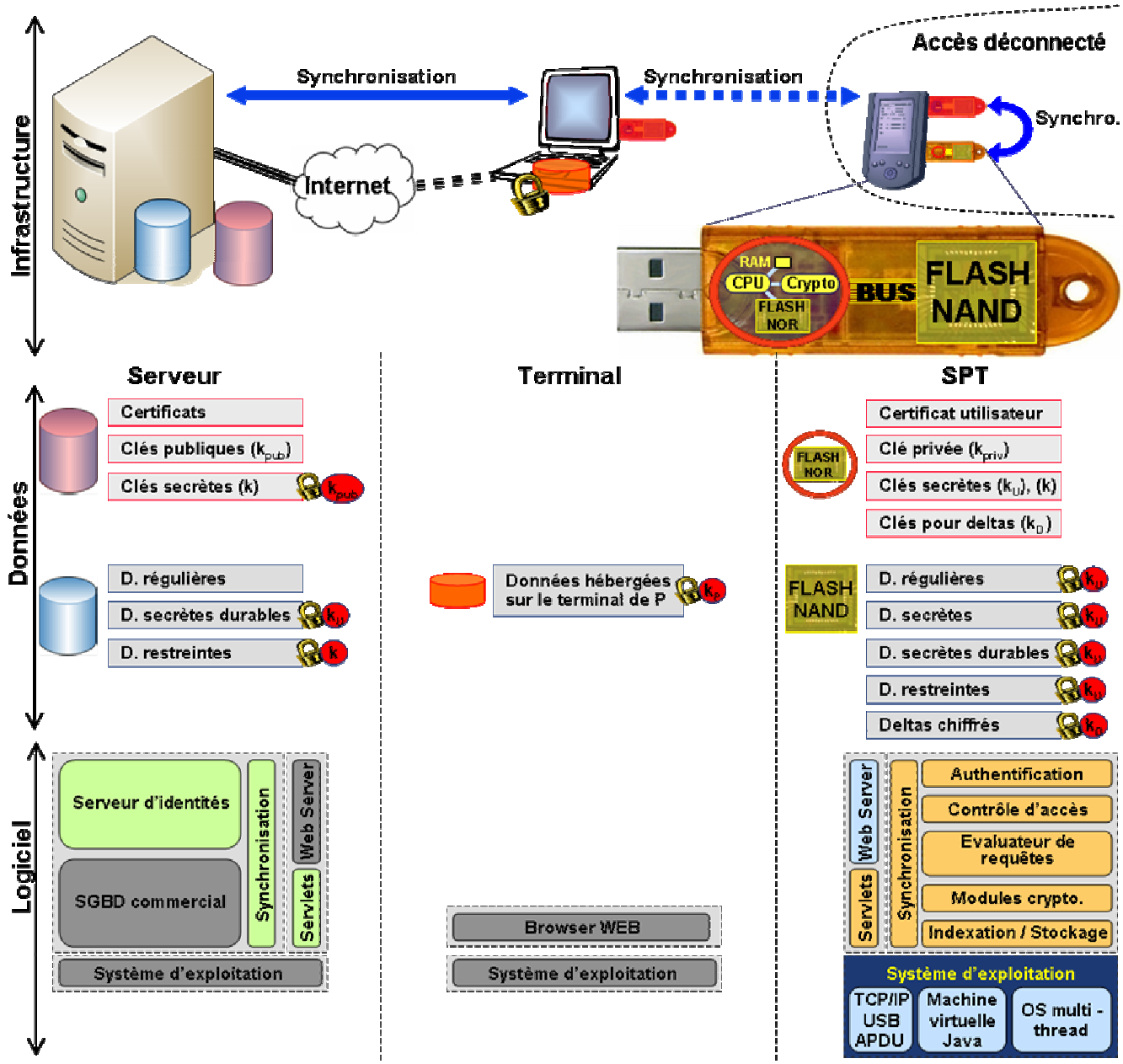

Figure 2 : Architecture fonctionnelle.

3.3 Composants logiciels

Le serveur central conserve une architecture classique ; dans le cas du projet il sera équipé d'un logiciel serveur Web, d'un SGBD chargé de stocker, indexer et restituer les données des dossiers personnels. Le serveur sera en outre doté des dispositifs de sécurité requis (firewall, chiffrement des données, etc.) pour assurer la sécurité des dossiers gérés.

L'interface utilisateur sera similaire que l'on se connecte au serveur central ou au serveur embarqué, via un terminal fixe ou mobile. Il s'agira d'une interface web connectant le serveur (central ou SPT) par des appels HTTP classiques.

Les composants logiciels embarqués dans le SPT sont : un système d'exploitation propriétaire, une machine virtuelle Java, un serveur Web et un SGBD :

- *Le système d'exploitation*. Il s'appuie sur un processeur 32 bits à haute performance disposant d'une mémoire de travail (RAM) de taille raisonnable (plusieurs dizaines de kilo-octets). La communication entre le SPT et le terminal répond à la fois à des contraintes de performance et à un objectif d'intégration la plus simple possible. Pour cela, la communication avec le mode extérieur se fait via un canal de communication USB supportant nativement (c'est à dire non traduit en termes d'APDU) le protocole Internet TCP/IP (et donc HTTP et HTTPS). Le système d'exploitation fournira également les services de base nécessaires à une architecture moderne de cartes de nouvelle génération s'appuyant sur un environnement Java Card, dont les spécifications fonctionnelles sont en cours d'élaboration et seront publiées en 2008.
- *La machine virtuelle et le serveur Web*. La machine virtuelle Java Card de nouvelle génération peut être vue comme équivalente à une machine virtuelle pour téléphones portables. Elle fournit le support du « multi-threading » pour les applications, un ramasse-miettes (« garbage collector ») automatique pour récupérer la mémoire occupée par les objets qui ne sont plus utilisés. Elle implémente également les services permettant d'établir des communications de types « sockets » clientes ou serveurs en TCP/IP. Deux modèles d'applications sont fournis par la plate-forme via deux « conteneurs » permettant de charger et d'exécuter des applications de types Web ou de type applets APDU. L'accès au dossier nomade dans le SPT pourra se faire au travers de services ou d'applications Web développés comme des Servlets ; Servlets qui elles-mêmes accéderont au données du dossier via une API Java au SGBD.
- *Le SGBD.* Afin de pouvoir gérer des dossiers nomades et sécurisés, le SPT intègre un véritable moteur de SGBD. Ce SGBD embarqué a pour objet le **stockage** de données personnelles dans la Flash NAND externe, leur **indexation** [ABB+07, PuS07] pour une recherche efficace, la journalisation des mises à jour pour assurer la propriété transactionnelle d'atomicité, l'évaluation de requêtes assertionnelles (i.e. à base de prédicats) et le **contrôle de droits d'accès** sophistiqués (également assertionnels comme dans toute base de données, par exemple, seules les données satisfaisant la qualification Q1 sont accessibles aux usagers satisfaisant la qualification Q2). Il est impératif que le contrôle d'accès soit embarqué dans le microcontrôleur sécurisé pour assurer sa résistance aux attaques lorsque le dossier est accédé via un terminal vulnérable ou hostile. Le contrôle d'accès s'appuyant sur le **gestionnaire de requêtes** pour évaluer les prédicats présents dans les règles de droits d'accès, et ce dernier s'appuyant à son tour sur le moteur de stockage et d'indexation, c'est l'intégralité du SGBD qui doit être embarqué dans la puce sécurisée [PBV+01]. Contrairement à la mémoire de stockage des cartes à puce traditionnelles, la mémoire FLASH NAND du SPT n'est pas protégée matériellement contre les attaques et doit donc l'être par des **méthodes**

cryptographiques. Ces méthodes (chiffrement, hachage, contrôle de version) doivent se faire avec une granularité et un coût compatible avec la grande quantité d'accès aléatoires aux données générées par l'évaluation de requêtes bases de données.

Enfin, un module de **synchronisation** est embarqué dans le SPT afin de réaliser la synchronisation du dossier embarqué avec la copie résidant sur le serveur central, soit directement lorsque le SPT se trouve dans un environnement connecté, soit via un autre SPT servant de canal de communication (cf. scénario en section 2.7).

4 Conclusion

Ce document présente les choix architecturaux réalisés dans le projet depuis son démarrage. Ces choix sont basés sur une analyse détaillée des besoins, guidée en particulier par l'expérimentation visée dans le contexte médico-social, sans pour autant faire de concession sur la généralité de l'approche. Nous pensons que cette nouvelle approche peut apporter de nouvelles solutions à la gestion de données personnelles et susciter de nouvelles pratiques, notamment concernant le partage et la conservation de données sensibles, répondant ainsi aux quatre dimensions identifiées du problème.

Les deux coordinations gérontologiques participant au projet, l'ALDS et COGITEY, offriront un environnement d'expérimentation idéal. Cette expérimentation est projetée courant 2009 dans le département des Yvelines avec une centaine de patients volontaires et 25 intervenants à domicile. A moyen terme, ce projet pourrait être étendu à d'autres populations vulnérables en situation de précarité ou de handicap. Plus généralement, nous espérons que cette expérimentation constituera une avancée dans le domaine des prestations à domicile (sujets âgés, handicap, etc.) et que la plate-forme mise en place pourra servir de base à d'autres initiatives départementales (par exemple, délivrance d'une carte à puce périnatalité permettant de suivre une femme et son enfant sur une période de temps déterminée dans le cadre des réseaux de périnatalité).

5 Bibliographie

[ABB+07] N. Anciaux, M. Benzine, L. Bouganim, P. Pucheral, D. Shasha, GhostDB: querying visible and hidden data without leaks, ACM SIGMOD Conference, 2007.

[ABF+04] R. Agrawal, R. J. Bayardo Jr., C. Faloutsos, J. Kiernan, R. Rantzau, R. Srikant, Auditing Compliance with a Hippocratic Database, Conference on Very Large Data Bases (VLDB), 2004.

[ACA05] ACAS, *Personnel data and record keeping*. Advisory Booklet. London, 2005. http://www.acas.org.uk/index.aspx?articleid=717

[AKSX02] R. Agrawal, J. Kiernan, R. Srikant, Y. Xu, Hippocratic Databases, Conference on Very Large Data Bases (VLDB), 2002.

[CLM+02] Cranor L., Langheinrich M., Marchiori M., Presler-Marshall M., and Reagle J. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification. W3C Recommendation, 2002.

[CSI06] Computer Security Institute, CSI/FBI Computer Crime and Security Survey, http://www.gocsi.com, 2006.

[EuD95] European Directive 95/46/EC, Protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal of the European Communities of 23 November 1995 No L. 281 p. 31

[HIPAA] Health Insurance Portability and Accountability Act of 1996, Public Law 104-191, 1996. http://aspe.hhs.gov/admnsimp/pl104191.htm

[IBM] IBM-Harris, IBM survey of consumer attitudes toward privacy in the United States, the United Kingdom and Germany, http://www.securitymanagement.com/library/ibm_priv.html.

[IC03] IPSOS Santé / CNAM, « Opinions et attitudes des médecins et des patients à l'égard du Dossier médical partagé », octobre 2003.

[LAT06] Los Angeles Times, "At Risk of Exposure", article published in June 26, 2006. On line : http://www.latimes.com/features/printedition/health/la-he-privacy26jun26,1,1971062.column?coll=la-headlines-pe-health&ctrack=1&cset=true

[MLS07] G. Miklau, B. N. Levine, P. Stahlberg, Securing history: Privacy and accountability in database systems, Conference on Innovative Data Systems Research (CIDR), 2007.

[PBV+01] P. Pucheral, L. Bouganim, P. Valduriez, C. Bobineau, "PicoDBMS: Scaling down Database Techniques for the Smartcard", Very Large Data Bases Journal (VLDBJ), Vol.10, n°2-3. October 2001. Extended version of the paper rewarded by the Best Paper Award of the Int. Conf. on Very Large Data Bases (VLDB'00).

[PuS07] P. Pucheral, S. Yin, 'System and Method of Managing Indexation of Flash Memory', *Dépôt par Gemalto et INRIA du brevet européen n° 07290567.2-*, 04/05/2007.

[West05] A. Westin, "Public Attitudes Toward Privacy and EHR Programs,", AHRQ Conference, Washington, 2005.

